

BAB II

TINJAUAN PUSTAKA

2.1 *Machine Learning*

Machine learning atau pembelajaran mesin merupakan penerapan kecerdasan buatan yang memberikan sistem kemampuan belajar secara otomatis dari sekumpulan data untuk melakukan tugas tertentu tanpa di program secara eksplisit (Batta, 2020). Sedangkan menurut (Rebala dkk., 2019), *machine learning* merupakan bidang ilmu komputer yang mempelajari algoritma dan teknik untuk menghasilkan sebuah solusi secara otomatis pada masalah kompleks yang sulit dilakukan dengan menggunakan metode pemrograman konvensional. Pada prosesnya algoritma pada *machine learning* akan mencari pola tertentu pada setiap data atau yang disebut dengan *dataset* untuk menghasilkan sebuah *rules* (aturan).

Secara umum, metode pada *machine learning* dibagi menjadi empat tipe berdasarkan cara pembelajarannya (Yu & He, 2019), yaitu:

a. *Supervised Learning*

Supervised learning merupakan metode pada *machine learning* yang memiliki fungsi dalam memetakan data masukan dan keluaran berdasarkan contoh yang tersedia atau yang disebut dataset. Dataset masukan terdiri dari data latih dan data uji, dimana data latih memiliki variabel keluaran yang perlu diprediksi atau diklasifikasi. Pada prosesnya metode ini akan mempelajari pola dari dataset latih dan menerapkannya ke setiap dataset uji untuk melakukan prediksi atau klasifikasi.

b. *Unsupervised Learning*

Unsupervised learning merupakan metode pada *machine learning* yang digunakan untuk melakukan pengelompokan data tanpa pengawasan, artinya dataset yang dikelola tidak mempunyai label tertentu. Pada prosesnya metode ini akan mengeksplorasi dan menemukan struktur dalam data untuk mengenali kelas data. Teknik yang digunakan pada *computer vision* yaitu *density estimation* (estimasi kepadatan) dan *dimensionality reduction* (reduksi dimensi) yang mengarah ke berbagai aspek pada *visual features* (fitur visual) (Caron dkk., 2019).

c. *Semi-supervised Learning*

Semi-supervised learning merupakan kombinasi metode *supervised learning* dan *unsupervised learning*, artinya data yang memiliki label dan tanpa label akan digunakan untuk melatih algoritma agar dapat melakukan proses pembelajaran sendiri. Metode ini digunakan untuk data yang memiliki label dengan jumlah yang sedikit dan data tanpa label dalam jumlah yang lebih banyak.

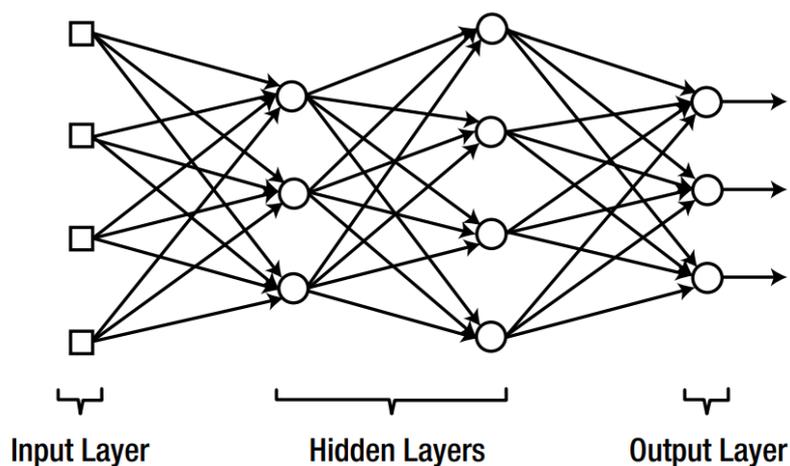
d. *Reinforcement Learning*

Reinforcement learning merupakan metode pada *machine learning* yang berhubungan dengan pengambilan keputusan secara berurutan untuk memperoleh hasil yang terbaik. Pada prosesnya metode ini akan mempelajari struktur data yang dapat memodifikasi atau memperoleh perilaku dan keterampilan baru secara bertahap untuk mendapatkan *reward* berdasarkan *trial and error experience* (pengalaman coba- coba) (François-lavet dkk., 2018).

2.2 Deep Learning

Deep learning atau pembelajaran mendalam merupakan sub-bidang yang spesifik dari *machine learning* dengan penerapan algoritma *artificial neural network* (jaringan syaraf tiruan) untuk mempelajari representasi data yang tidak terstruktur dalam jumlah yang besar (Pramesty, 2018). Pada *deep learning* lapisan representasi dipelajari melalui model yang disebut *neural network* (jaringan syaraf) dalam lapisan literal yang ditumpuk satu sama lain secara terstruktur.

Terdapat tiga lapisan utama pada *deep learning*, diantaranya *input layer*, *hidden layer* dan *output layer*. Berikut merupakan ilustrasi lapisan arsitektur jaringan syaraf pada *deep learning*.



Gambar 2. 1 Struktur *Neural Network* pada *Deep Learning*
(Sumber : Nima & Shila, 2020)

Arsitektur *deep learning* pada Gambar 2. 1 disebut sebagai *multi layer perceptron*, yang memiliki 4 *node neuron* pada *input layer* (lapisan masukan) berfungsi sebagai jalur untuk mentransmisikan sinyal masukan ke *node* berikutnya, 2 *hidden layer* (lapisan tersembunyi) yang merupakan perantara antara *input layer* dengan *output layer* dan berfungsi untuk menambah instruksi logika, serta 3 *node*

neuron pada *output layer* (lapisan keluaran) yang menjadi hasil akhir dari *neural network*.

Jumlah *neuron* yang diterapkan pada sebuah lapisan tersembunyi sangat mempengaruhi performa *deep learning*. Jika jumlah *node neuron* yang diterapkan pada arsitektur *deep learning* terlalu sedikit maka dapat menyebabkan model menjadi *underfitting*. Jika jumlah *node neuron* yang diterapkan terlalu banyak maka dapat menyebabkan model menjadi *overfitting*.

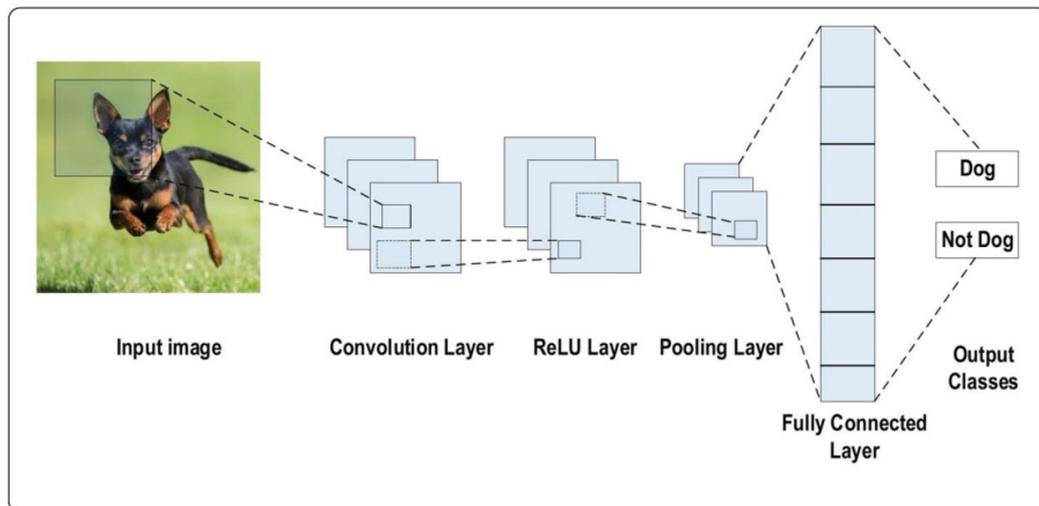
Penelitian ini menerapkan teknik *deep learning* pada metode *machine learning* yaitu *supervised learning* untuk melakukan klasifikasi HAR pada video, dengan menggunakan kombinasi algoritma CNN dan LSTM.

2.3 Convolutional Neural Network

Convolutional Neural Network (CNN) merupakan sebuah arsitektur pada *deep learning* yang memiliki banyak lapisan representasi atau disebut *multi layer perception* untuk mengolah data dua dimensi. CNN memiliki struktur *feed-forward* yang dalam dan memiliki kemampuan generalisasi yang lebih dibandingkan jaringan lain dengan lapisan konvensional *fully-connected*, yang dapat melakukan proses pembelajaran abstrak dengan fitur objek terutama pada data spasial (Ghosh dkk., 2019) dan dapat mengidentifikasinya secara lebih efisien, karena memiliki karakteristik representasi data pada transformasi nonlinier serta penerapan fungsi nonlinier (Firmansyah, 2020).

Secara teknis, model CNN terdiri dari serangkaian lapisan pemrosesan terbatas yang dapat melakukan proses pembelajaran dengan berbagai fitur data

masuk dengan beberapa tingkatan abstraksi. Lapisan inisiator akan melakukan proses pembelajaran dan ekstraksi fitur tingkat tinggi dengan abstraksi yang lebih rendah, dan lapisan yang lebih dalam akan mempelajari dan mengekstrak fitur tingkat rendah dengan abstraksi yang lebih tinggi.



Gambar 2. 2 Arsitektur Algoritma CNN

(Sumber : Alzubaidi dkk., 2021)

2.3.1 Lapisan *Convolutional Neural Network*

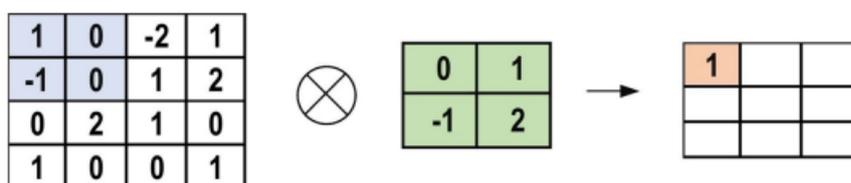
Berdasarkan Gambar 2. 2, CNN memiliki 3 lapisan yang merupakan struktur *neural network* pada *deep learning*, yaitu lapisan masukan, lapisan keluaran, dan lapisan tersembunyi, dimana pada lapisan tersembunyi terdapat 4 macam lapisan utama pada arsitektur CNN yaitu *convolutional layer*, *relu layer*, *pooling layer*, dan *fully connected layer*. Berikut merupakan penjelasan dari lapisan utama pada arsitektur CNN:

a. *Convolution Layer*

Dalam arsitektur CNN, komponen yang paling signifikan adalah *convolutional layer* atau lapisan konvolusional. Pada lapisan ini, terdiri dari

kumpulan filter konvolusional yang disebut kernel untuk melakukan proses komputasi dengan mengaplikasikan sebuah fungsi pada fungsi keluaran lain secara berulang. Gambar masukan, dinyatakan sebagai metrik *N-dimensi* yang diproses dengan filter ini untuk menghasilkan peta fitur keluaran.

Kernel dapat didefinisikan sebagai angka atau nilai diskrit pada sebuah kotak. Pada Gambar 2. 3, setiap nilai pada kotak hijau disebut dengan bobot *kernel*. Terdapat nomor acak yang bertindak sebagai bobot *kernel* pada awal proses pelatihan. Selanjutnya, bobot tersebut akan disesuaikan pada setiap proses pelatihan untuk mengekstraksi fitur-fitur yang penting. Nilai yang dihasilkan (ditandai dengan warna oranye) merupakan nilai masuk pada peta fitur keluaran.



Gambar 2. 3 Ilustrasi Operasi Konvolusi

(Sumber : Alzubaidi dkk., 2021)

Terdapat dua parameter penting pada lapisan konvolusi yaitu *stride* dan *padding*. *Stride* digunakan filter untuk menggeser gambar. *Stride* bernilai satu merupakan ukuran langkah pada setiap lokasi secara *vertical* dan *horizontal* yang diterapkan pada *kernel*. *Stride* bernilai dua akan membuat ukuran langkah pada lapisan konvolusi dengan setengah lebar dan tinggi gambar (Firmansyah, 2020). Di sisi lain, parameter *padding* memiliki peran penting untuk menentukan informasi ukuran batas yang terkait dengan gambar masukan dan ukuran peta fitur keluaran menjadi akan meningkat.

b. *Rectified Linear Unit Layer (ReLU)*

ReLU merupakan fungsi aktivasi yang paling umum digunakan dalam konteks CNN. Representasi matematis dari fungsi ReLU adalah $f(x) = \max(0, x)$. *ReLU* berfungsi untuk mengubah seluruh nilai input menjadi angka positif. Beban komputasi yang lebih rendah adalah keuntungan utama pada fungsi *ReLU*.

c. *Pooling Layer*

Pooling layer memiliki tugas utama untuk mengecilkan peta fitur yang berukuran besar menjadi lebih kecil. *Pooling layer* berfungsi untuk mempertahankan sebagian besar informasi atau fitur disetiap langkah penyautan. Terdapat beberapa jenis metode *pooling* yang tersedia untuk digunakan pada lapisan *pooling*, diantaranya *tree pooling*, *gated pooling*, *average pooling*, *min pooling*, *max pooling*, *global average pooling*, dan *global max pooling*.

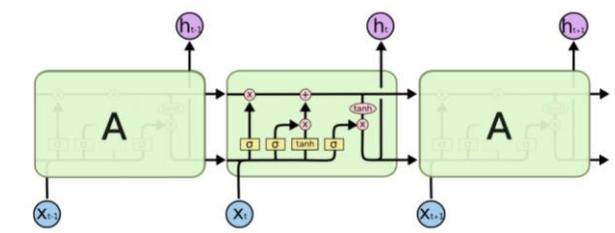
Penerapan parameter *pooling layer* dapat mempengaruhi kinerja CNN secara keseluruhan menjadi menurun, karena lapisan ini membantu CNN untuk fokus dalam menentukan ketersediaan fitur pada gambar masukan yang membuat terlewatnya informasi yang relevan.

d. *Fully-Connected Layer*

Fully-Connected layer merupakan proses akhir pada arsitektur CNN untuk melakukan proses klasifikasi. Di dalam lapisan ini, setiap *neuron* terhubung ke semua *neuron* dari lapisan sebelumnya, yang disebut pendekatan *Fully Connected* (FC). Masukan dari lapisan FC berasal dari *pooling* atau lapisan konvolusi terakhir berupa vektor, yang dihasilkan dari peta fitur.

2.4 Long Short-Term Memory (LSTM)

LSTM merupakan jenis model atau struktur pada data sekuensial yang dikembangkan oleh (Hochreiter & Schmidhuber, 1997) untuk pembaharuan pada model RNN. Pada arsitektur LSTM menggunakan kombinasi khusus dari unit tersembunyi, produk elemen, dan jumlah antar unit untuk menerapkan gerbang yang dapat mengontrol sel memori (Apaydin dkk., 2020). Sel memori pada LSTM dirancang untuk menyimpan informasi tanpa modifikasi untuk waktu yang lama. Fitur terbaik yang dimiliki LSTM adalah kemampuannya untuk mempelajari dependensi jangka panjang, yang tidak dapat dilakukan pada model RNN. Tujuan pembaharuan model LSTM adalah untuk mengatasi masalah *vanishing gradient* yang sering terjadi pada model RNN. Gambar 2. 4 merupakan ilustrasi sel memori pada model LSTM pada *single layer*. LSTM memiliki komponen utama yang disebut *gates units*. *Gates units* tersebut terdiri dari *forget gates*, *input gates*, *cell gates*, dan *output gates* (Wildan dkk., 2018).



Gambar 2. 4 Arsitektur Sel Memori pada LSTM

(Sumber : Wildan dkk., 2018)

Forget gates (f_t) merupakan langkah awal yang dilakukan pada model LSTM untuk mengolah dan menyaring setiap masukan pada sel memori menggunakan fungsi sigmoid. Data (f_t) memiliki hasil antara 0 dan 1. Jika data (f_t)

menghasilkan 1 maka data akan disimpan dan jika menghasilkan 0 data akan dibuang.

Berikut merupakan rumus perhitungan nilai *forget gates*:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (2.1)$$

Setelah proses penyaringan selesai maka informasi akan diolah dengan menentukan informasi yang akan diperbarui melalui komponen *input gates* (i_t) menggunakan fungsi aktivasi sigmoid. Pada langkah ini akan membuat nilai vektor baru menggunakan fungsi aktivasi tanh yang akan disimpan pada sel memori.

Berikut merupakan rumus perhitungan nilai *input gates* :

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2.2)$$

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (2.3)$$

Pada *cell gates* (c_t) akan merubah nilai pada sel memori melalui proses penggabungan nilai yang dihasilkan pada (f_t) dan (i_t). Berikut merupakan rumus perhitungan nilai *cell gates*:

$$c_t = f_t * c_{t-1} + i_t * \tilde{c}_t \quad (2.4)$$

Output gates (o_t) merupakan langkah terakhir yang dilakukan pada model LSTM. Pada proses ini, akan menentukan nilai keluaran pada sel memori menggunakan fungsi aktivasi *sigmoid* dan menempatkan ke sel memori menggunakan fungsi aktivasi tanh. Nilai keluaran yang dihasilkan dari kedua fungsi aktivasi tersebut dilakukan perkalian untuk menghasilkan nilai klasifikasi dari model LSTM. Berikut merupakan perhitungan nilai *Output gates*:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (2.5)$$

$$h_t = o_t * \tanh(c_t) \quad (2.6)$$

2.5 *Human Activity Recognition (HAR)*



Gambar 2. 5 Aktivitas Olahraga Dataset UCF50

(Sumber : Reddy & Shah, 2013)

HAR merupakan teknik untuk mengenali berbagai aktivitas manusia melalui sensor eksternal seperti sensor inersia atau video. HAR berbasis video merupakan bidang penelitian yang menarik dan menantang di era ini terutama dalam mengidentifikasi dan mengenali tindakan manusia dibidang pengawasan keamanan berbasis video, bantuan medis, serta realitas virtual. HAR merupakan kunci untuk lebih memahami berbagai semantik yang terkandung dalam video. Metodologi dasar pengenalan aktivitas melibatkan ekstraksi fitur aktivitas, pemodelan, dan teknik pengenalan.

HAR berbasis video dapat dibagi menjadi dua kategori menurut fitur gerak, yaitu *marker based* (berbasis penanda) dan *vision-based* (berbasis visi). Pada sistem *motion capture* berbasis penanda optic, banyak digunakan karena menawarkan keuntungan menangkap gerakan manusia yang kompleks secara akurat, namun memiliki kekurangan yaitu sensor optik harus terpasang pada tubuh dan membutuhkan pengaturan kamera ganda. Gambar 2. 5 merupakan aktivitas manusia dibidang olahraga dengan pendekatan berbasis visi yang mendapatkan lebih banyak

perhatian akhir-akhir ini karena tidak adanya penanda yang selalu dipakai untuk mendeteksi aktivitas (Park dkk., 2016).

2.6 Tools

2.6.1 Python

Python adalah bahasa pemrograman komputer yang sering digunakan untuk membangun situs web dan perangkat lunak, mengotomatisasi tugas, dan melakukan analisis data. *Python* merupakan *general-purpose language*, artinya dapat digunakan untuk membuat berbagai program pada domain yang berbeda dan tidak khusus pada masalah tertentu.

2.6.2 Tensorflow

TensorFlow adalah *library open-source* yang dibuat oleh tim *Google Brain*, untuk melakukan proses komputasi numerik dan pembelajaran mesin pada skala yang besar. *TensorFlow* menggabungkan banyak model pembelajaran mesin dan pembelajaran mendalam yang mendukung proses prediksi dalam skala besar, dengan model yang sama yang digunakan untuk melakukan proses pelatihan. *TensorFlow* menggunakan *Python* untuk menyediakan *API front-end* yang nyaman dalam membangun aplikasi dengan *framework* yang tersedia.

2.7 Penelitian Terkait dan Keterbaruan Penelitian

Salah satu metode metode pengumpulan data yang dilakukan yaitu dengan studi literatur. Pada studi literatur ini yaitu dengan mengumpulkan beberapa penelitian terdahulu dan melakukan pendalaman materi mengenai model *machine learning* pada HAR dengan metode CNN dan LSTM yang nantinya akan menjadi acuan dan perbandingan dalam melakukan penelitian ini. Berikut merupakan referensi penelitian sejenis yang dijelaskan ke dalam bentuk Tabel 2. 1.

Tabel 2. 1 Referensi Penelitian Sejenis

No.	Judul, Sumber, Dan Tahun Penelitian	Peneliti	Metode	Fokus Penelitian	Hasil
1.	Human action recognition using attention based LSTM network with dilated CNN features. (Elsevier, 2021)	Khan Muhammad, Mustaqeem, Amin Ullah, Ali Shariq Imran,	Bi-LSTM, DCNN	Penentuan perbedaan fitur, seperti perbedaan visual dengan temporal pada teknik pengenalan tindakan menggunakan bobot pra-terlatih dari arsitektur	Peneliti mengusulkan algoritma <i>bi-directional long short-term memory</i> (Bi-LSTM) berbasis mekanisme atensi dengan <i>dilated convolutional neural network</i> (DCNN) yang secara

		Muhammad Sajjad, Mustafa Servet Kiran, Giovanna Sannino, Victor Hugo C. de Albuquerque.		pembelajaran mesin untuk representasi visual <i>frame</i> video pada tahap pelatihan.	selektif berfokus pada fitur dalam masukan <i>frame</i> yang efektif untuk mengenali berbagai tindakan manusia dalam video.
2.	Human action recognition using convolutional LSTM and fully-connected	Zufan Zhang, Zongming Lv, Chenquan Gan, Qingyi Zhu.	Conv-LSTM, FC-LSTM	Memfasilitasi pengenalan tindakan dengan arsitektur jaringan perhatian ganda STDAN, yang menggabungkan	Pada hasil eksperimen menunjukkan bahwa STDAN yang diusulkan memiliki kinerja pengenalan yang lebih baik daripada metode yang ada.

	LSTM with different attentions. (Elsevier, 2020)			<i>Convolutional LSTM</i> dan <i>Fully-Connected LSTM</i> untuk menangani fitur sekuensial berdurasi panjang dengan informasi konteks temporal yang berbeda dan pencegahan terjadinya <i>overfitting</i> .	
3.	Action recognition using three dimension convolution and long short term memory. (IEEE, 2017)	Yu-Cheng Liu, Jian-Jiun Ding, Yao-Jen Chang, Chien-Yao Wang, and Jia-Ching Wang.	3D-CNN, LSTM	Penggunaan algoritma CNN 3D dan LSTM dalam identifikasi aktivitas manusia pada video yang mengandung informasi temporal dan pencegahan permasalahan penting pada <i>overfitting</i> .	Peneliti mengusulkan arsitektur yang merupakan kombinasi dari CNN 3D dan LSTM diusulkan untuk identifikasi aktivitas manusia. Pada proses simulasi menunjukkan bahwa algoritma yang diusulkan memiliki kinerja

					yang lebih baik untuk identifikasi aktivitas manusia.
4.	Improved two-stream model for human. (EURASIP Journal on Image, 2020)	Yuxuan Zhao, Ka Lok Man, Jeremy Smith, Kamran Siddique, Sheng-Uei Guan.	CNN, LSTM	Penerapan struktur <i>two-stream</i> pada model berbasis LSTM dalam aliran spasialnya untuk mengekstrak fitur spasial dan temporal pada <i>RGB frame</i> .	Metode yang diusulkan mengungguli metode CNN <i>two-stream</i> tradisional dengan akurasi minimal 3%. Untuk aliran spasial dan temporal, model yang diusulkan juga mencapai akurasi pengenalan yang lebih tinggi.
5.	3D-CNN-Based Fused Feature Maps with LSTM Applied to Action Recognition.	Sheeraz Arif, Jing Wang, Tehseen Ul	3D-CNN, LSTM	Pemanfaatan fitur dalam <i>frame-level</i> dari 3D-CNN dan memprosesnya melalui LSTM. Pada prosesnya metode fusi	Metode yang diusulkan mampu membuktikan keefektifan dan kepraktisan dalam melakukan

	(Future Internet, 2019)	Hassan, Zesong Fei.		berbobot linier dirancang untuk menggabungkan peta fitur spasial dan temporal secara efektif.	proses <i>machine learning</i> pada HAR.
6.	3D-CNN-Based Fused Feature Maps with LSTM Applied to Action Recognition. (ICPR, 2018)	Wenchao Xu, Yuxin Pang, Yanqin Yang, Yanbo Liu.	CNN	Terdapat hasil identifikasi yang tidak diketahui dan kurangnya tingkat akurasi yang menyebabkan kinerja identifikasi rendah.	Peneliti mengusulkan algoritma CNN menggunakan data akselerometer <i>triaxial</i> dan mencapai akurasi pengenalan yang tinggi dengan penggunaan komputasi yang rendah untuk pra-proses data, yang mengungguli metode berbasis SVM dalam klasifikasi multi aktivitas manusia.

7.	Human Activity Recognition Based On Optimal Skeleton Joints Using Convolutional Neural Network. (ICCSIT, 2018)	Nor Surayahani Suriani, Siti Noor Fatihah Ahmad, Mohd Norzali Mohd, Mohd Razali Tomari, Wan Nurshazwani Wan Zakaria.	CNN	Peneliti mengusulkan sambungan informatif kerangka yang dievaluasi dengan rumus entropi shannon dan hanya memilih titik paling signifikan untuk klasifikasi lebih lanjut menggunakan CNN.	Metode yang diusulkan mencapai 94,16% pada kumpulan data CAD-60.
8.	A CNN-LSTM Approach to Human Activity Recognition. (IEEE, 2020)	Ronald Mutegeki, Dong Seog Han.	CNN, LSTM	Pada proses multi-kelas klasifikasi terdapat kesalahan prediksi karena banyaknya parameter masukan.	Pendekatan CNN-LSTM yang dilakukan tidak hanya meningkatkan akurasi prediksi aktivitas manusia, tetapi juga

					mengurangi kompleksitas model dan menghilangkan kebutuhan akan rekayasa fitur tingkat lanjut.
9.	Leveraging CNN and Transfer Learning for Vision-based Human Activity Recognition. (IEEE, 2019)	Samundra Deep, Xi Zheng.	CNN	Peneliti menerapkan algoritma CNN dan menggunakan <i>transfer learning</i> pada proses pelatihan.	Metode yang dilakukan menghasilkan akurasi 96,95% menggunakan VGG-16 dan mengungguli dalam hal ekstraksi fitur.
10.	A Depth Camera-based Human Activity Recognition via Deep Learning Recurrent Neural Network for	S. U. Park, J. H. Park, M. A. Al-masni, M. A. Al-antari,	RNN	Penerapan algoritma RNN pada penggunaan sudut sendi dari beberapa sendi tubuh yang berubah dalam waktu yang	Peneliti mengusulkan model RNN yang dapat menghasilkan tingkat akurasi rata-rata sebesar 99.55%. Tingkat akurasi RNN yang diusulkan lebih tinggi

	Health and Social Care Services. (Elsevier, 2016)	Md. Z. Uddin, T.-S.Kim		direpresentasikan sebagai matriks fitur spatiotemporal	dibandingkan dengan model konvensional HMM dan DBN pada HAR.
11.	Long-term Recurrent Convolutional Networks for Visual Recognition and Description. (IEEE, 2016)	Jeff Donahue, Lisa Anne Hendricks, Marcus Rohrbach, Subhashini Venugopalan, Sergio Guadarrama, Kate Saenko, Trevor Darrell	LRCN	Penelitian ini berfokus pada perancangan model <i>machine learning</i> yaitu penerapan kelas arsitektur konvolusi berulang yang dapat dilatih secara <i>end to end</i> dan sesuai pada tugas pemahaman visual skala besar untuk pengenalan aktivitas, teks gambar, dan deskripsi video.	Penelitian yang diusulkan mampu menunjukkan bahwa model tersebut memiliki keunggulan untuk pengenalan atau proses generasi yang ditentukan dan dapat dioptimalkan secara terpisah.

2.7.1 Matriks Penelitian dan *State Of The Art*

Matriks penelitian merupakan gambaran keseluruhan metode penelitian yang digunakan pada penelitian terdahulu dan yang dilakukan. Terdapat beberapa perbedaan dan persamaan pada penggunaan metode antara penelitian terdahulu dengan penelitian yang dilakukan. Berikut merupakan matriks penelitian yang disajikan pada Tabel 2. 2:

Tabel 2. 2 Matriks Penelitian

No.	Penulis	Penggunaan Metode							
		CNN	DCNN	RNN	3D-CNN	Bi-LSTM	FC-LSTM	Conv-LSTM	LSTM
1.	Muhammad dkk., 2021		✓			✓			
2.	Zhang dkk., 2020						✓	✓	
3.	Liu dkk., 2017				✓				✓
4.	Zhao dkk., 2020	✓							✓
5.	Arif dkk., 2019				✓				✓
6.	Xu dkk., 2018	✓							
7.	Surayahani dkk., 2018	✓							
8.	Mutegeki & Han dkk., 2020	✓							✓
9.	Deep & Zheng dkk., 2019	✓							
10.	Park dkk., 2016			✓					
11.	Donahue dkk., 2016	✓							✓
12.	Putra dkk., 2022	✓							✓

Penelitian yang dilakukan bertujuan untuk mengetahui tingkat parameter kualitas dalam perancangan arsitektur model *machine learning* dan menganalisa permasalahan yang sering terjadi dalam melakukan proses klasifikasi pada HAR.

Terdapat penelitian yang berfokus pada penggunaan fitur spasial dan temporal yang digunakan untuk memproses data gambar dalam visual video pada HAR seperti yang dilakukan oleh (Muhammad dkk., 2021) dan (Arif dkk., 2019).

Kemudia dilanjut dengan penelitian mengenai HAR yang berfokus pada proses multi-kelas klasifikasi dalam peningkatan akurasi dengan biaya komputasi rendah (Xu dkk., 2018) dan pengurangan kompleksitas model dengan proses eliminasi yang dibutuhkan untuk teknik fitur tingkat lanjut (Mutegeki & Han, 2020).

Dalam mengatasi terjadinya *overfitting* pada proses klasifikasi HAR sudah beberapa kali dibahas dalam penelitian yang dilakukan oleh (Liu dkk., 2017), (Zhao dkk., 2020), dan (Zhang dkk., 2020) yaitu dengan memperhatikan penerapan *dropout layer*.

Penelitian yang akan dilakukan menggunakan pendekatan arsitektur CNN yang dikombinasi dengan LSTM, serta penerapan optimasi adam yang umumnya lebih baik dibandingkan algoritma optimasi lainnya dalam melakukan proses dengan jumlah data yang banyak dan memiliki waktu komputasi serta penggunaan memori yang lebih efisien.

2.7.2 Relevansi Penelitian

Hubungan antara penelitian terkait dengan penelitian yang akan dilakukan tertera pada Tabel 2. 3.

Tabel 2. 3 Relevansi Penelitian Terkait dengan Penelitian yang Dilakukan

Penulis	Jeff Donahue (2016)	Rama Edwinda Putra (2022)
Judul	Long-term Recurrent Convolutional Networks for Visual Recognition and Description	Penerapan Algoritma Convolutional Neural Network dan Long Short-Term Memory Pada Human Activity Recognition Berbasis Pengolahan Visual Pada Video
Metode	CNN - LSTM	CNN - LSTM
Struktur Model	<i>CNN base of CaffeNet with pretrained ImageNet + 1 LSTM layer</i>	<i>4 convolution layer + 4 pooling layer + 1 flatten layer + 5 timedistributed layer + 1 fc layer + 1 LSTM layer</i>